

Online Learning from User Interactions through Interventions

CS 7792 - Fall 2016

Thorsten Joachims

Department of Computer Science & Department of Information Science
Cornell University

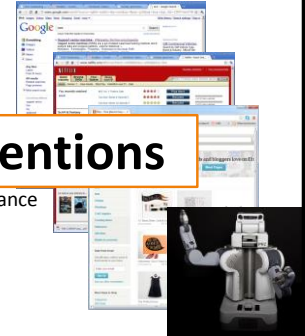


Y. Yue, J. Broder, R. Kleinberg, T. Joachims. The K-armed Dueling Bandits Problem. In COLT, 2009.
P. Shivaswamy, T. Joachims. Online Structured Prediction via Coactive Learning, ICML, 2012.

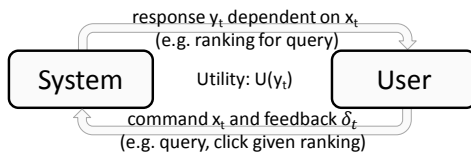
Interactive Learning Systems

- Examples
 - Search engines
 - Entertainment media
 - E-commerce
 - Smartphones
- Learning
 - Gathering and maintenance of knowledge
 - Measure and optimize performance
 - Personalization

Interventions

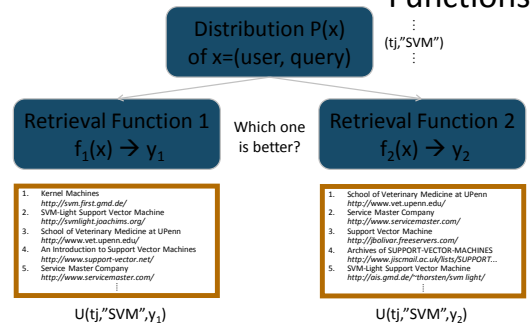


Interactive Learning System



- Information Elicitation from the User
 - Via generative behavioral model
 - Via information-elicitation interventions
- Online Learning with Interventions
 - Dueling Bandits: Algorithm-driven exploration
 - Coactive Learning: User-driven exploration

Decide between two Ranking Functions

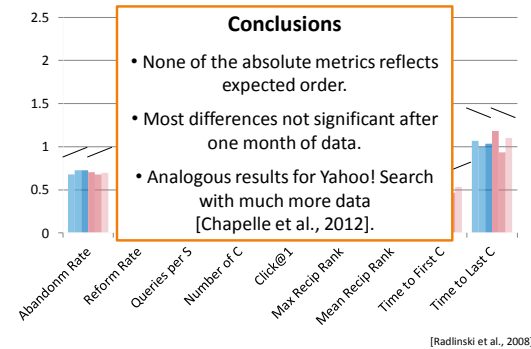


Measuring Utility

Name	Description	Aggregation	Hypothesized Change with Decreased Quality
Abandonment Rate	% of queries with no click	N/A	Increase
Reformulation Rate	% of queries that are followed by reformulation	N/A	Increase
Queries per Session	Session = no interruption of more than 30 minutes	Mean	Increase
Clicks per Query	Number of clicks	Mean	Decrease
Click@1	% of queries with clicks at position 1	N/A	Decrease
Max Reciprocal Rank*	1/rank for highest click	Mean	Decrease
Mean Reciprocal Rank*	Mean of 1/rank for all clicks	Mean	Decrease
Time to First Click*	Seconds before first click	Median	Increase
Time to Last Click*	Seconds before final click	Median	Decrease

(* only queries with at least one click count)

Arxiv.org: Results



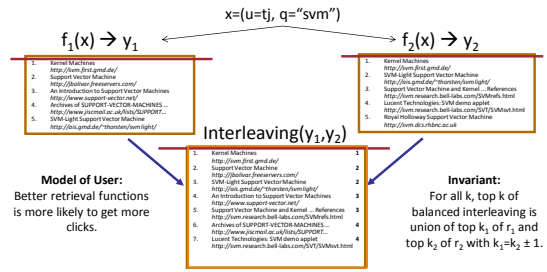
[Radlinski et al., 2008]

A Model of how Users Click in Search

- Model of clicking:
 - Users explore ranking to position k
 - Users click on most relevant (looking) links in top k
 - Users stop clicking when time budget up or other action more promising (e.g. reformulation)
 - Empirically supported by [Granka et al., 2004]



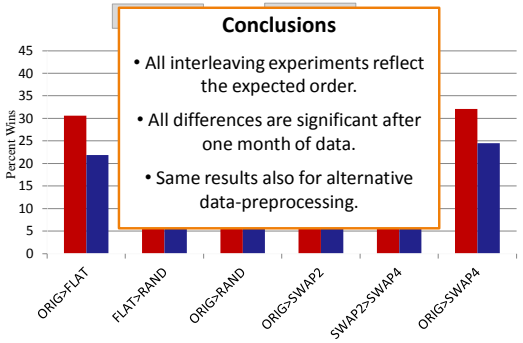
Balanced Interleaving



Interpretation: $(y_1 \succ y_2) \Leftrightarrow \text{clicks}(\text{top}_k(y_1)) > \text{clicks}(\text{top}_k(y_2))$
 → see also [Radlinski, Craswell, 2012] [Hofmann, 2012]

[Joachims, 2001][Radlinski et al., 2008]

Arxiv.org: Interleaving Results

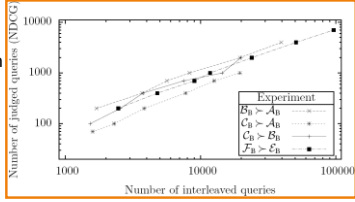


Yahoo and Bing: Interleaving Results

- Yahoo Web Search [Chapelle et al., 2012]
 - Four retrieval functions (i.e. 6 paired comparisons)
 - Balanced Interleaving
 - All paired comparisons consistent with ordering by NDCG.
- Bing Web Search [Radlinski & Craswell, 2010]
 - Five retrieval function pairs
 - Team-Game Interleaving
 - Consistent with ordering by NDCG when NDCG significant.

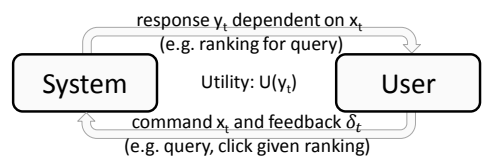
Efficiency: Interleaving vs. Explicit

- Bing Web Search
 - 4 retrieval function pairs
 - ~12k manually judged queries
 - ~200k interleaved queries
 - Experiment
 - p = probability that NDCG is correct on subsample of size y
 - x = number of queries needed to reach same p -value with interleaving
- Ten interleaved queries are equivalent to one manually judged query.



[Radlinski & Craswell, 2010]

Interactive Learning System



- Information Elicitation from the User
 - Via generative behavioral model
 - Via information-elicitation interventions ✓
- Online Learning with Interventions
 - Dueling Bandits: Algorithm-driven exploration
 - Coactive Learning: User-driven exploration

Learning on Operational System

- Example: 4 retrieval functions: $A > B \gg C > D$
 - 10 possible pairs for interactive experiment
 - (A,B) → low cost to user
 - (A,C) → medium cost to user
 - (C,D) → high cost to user
 - (A,A) → zero cost to user
 - ...
- Minimizing Regret
 - Don't present "bad" pairs more often than necessary
 - Trade off (long term) informativeness and (short term) cost
 - Definition: Probability of (f_t, f'_t) losing against the best f^*

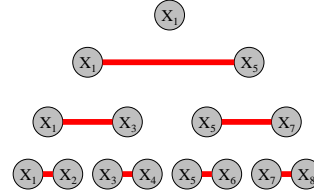
$$R(A) = \sum_{t=1}^T [P(f^* \succ f_t) - 0.5] + [P(f^* \succ f'_t) - 0.5]$$

→ Dueling Bandits Problem

[Yue, Broder, Kleinberg, Joachims, 2010]

First Thought: Tournament

- Noisy Sorting/Max Algorithms:
 - [Feige et al.]: Triangle Tournament Heap $O(n/\epsilon^2 \log(1/\delta))$ with prob $1-\delta$
 - [Adler et al., Karp & Kleinberg]: optimal under weaker assumptions



Algorithm: Interleaved Filter 2

Algorithm

InterleavedFilter1($T, W = \{f_1, \dots, f_k\}$)

- Pick random f' from W
- $\delta = 1/(TK^2)$
- WHILE $|W| > 1$
 - FOR $b \in W$ DO
 - » duel(f', b)
 - » update P_t
 - $t = t + 1$
 - $c_t = (\log(1/\delta)/t)^{0.5}$
 - Remove all f from W with $P_t < 0.5 - c_t$ [WORSE WITH PROB $1-\delta$]
 - IF there exists f'' with $P_{t'} > 0.5 + c_t$ [BETTER WITH PROB $1-\delta$]
 - » Remove f' from W
 - » Remove all f from W that are empirically inferior to f'
 - » $f' = f''$; $t = 0$
- UNTIL T : duel(f', f')

f_1	f_2	$f' = f_3$	f_4	f_5
0/0	0/0		0/0	0/0
f_1	f_2	$f' = f_3$	f_4	f_5
8/2	7/3		4/6	3/0
f_1	f_2	$f' = f_3$	f_4	
13/2	11/4	X	X	XX
$f' = f_1$	f_2		f_4	
0/0	0/0	XX	XX	XX

Related Algorithms: [Hofmann, Whiteson, Rijke, 2011] [Yue, Joachims, 2009] [Yue, Joachims, 2011]

[Yue et al., 2009]

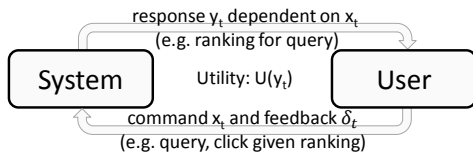
Assumptions

- Preference Relation: $f_i \succ f_j \Leftrightarrow P(f_i \succ f_j) = 0.5 + \epsilon_{i,j} > 0.5$
- Weak Stochastic Transitivity: $f_i \succ f_j$ and $f_j \succ f_k \rightarrow f_i \succ f_k$

Theorem: IF2 incurs expected average regret bounded by

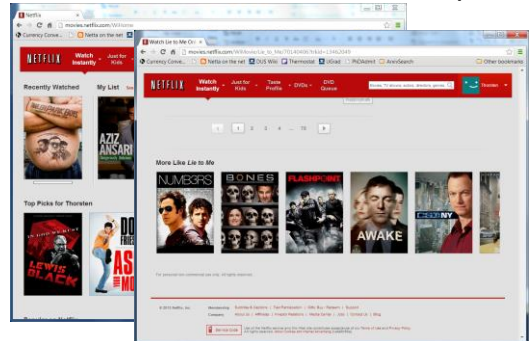
- Stochastic Triangle Inequality: $f_i \succ f_j \succ f_k \rightarrow \epsilon_{i,k} \leq \epsilon_{i,j} + \epsilon_{j,k}$
- $\epsilon_{1,2} = 0.01$ and $\epsilon_{2,3} = 0.01 \rightarrow \epsilon_{1,3} \leq 0.02$
- ϵ -Winner exists: $\epsilon = \max_i \{P(f_1 \succ f_i) - 0.5\} = \epsilon_{1,2} > 0$

Interactive Learning System

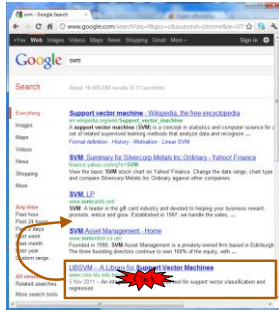


- Information Elicitation from the User
 - Via generative behavioral model
 - Via information-elicitation interventions ✓
- Online Learning with Interventions
 - Dueling Bandits: Algorithm-driven exploration ✓
 - Coactive Learning: User-driven exploration

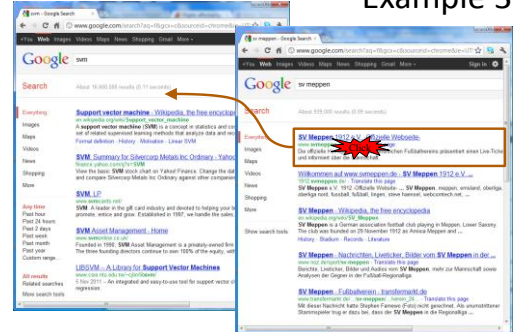
Who does the exploring? Example 1



Who does the exploring? Example 2

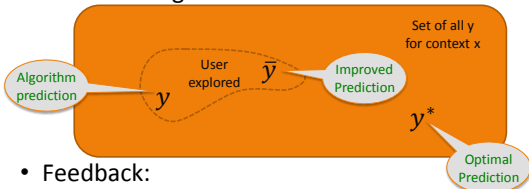


Who does the exploring? Example 3



Coactive Feedback Model

- Interaction: given x



- Feedback:

- Improved prediction \tilde{y}_t
 $U(\tilde{y}_t | x_t) > U(y_t | x_t)$
- Supervised learning: optimal prediction y_t^*
 $y_t^* = \operatorname{argmax}_y U(y | x_t)$

Machine Translation

X_t We propose Coactive Learning as a model of interaction between a learning system and a human user, where both have the common goal of providing results of maximum utility to the user.

Y_t Wir schlagen vor, koaktive Learning als ein Modell der Wechselwirkung zwischen einem Lernsystem und menschlichen Benutzer, wobei sowohl die gemeinsame Ziel, die Ergebnisse der maximalen Nutzen für den Benutzer.

\tilde{Y}_t Wir schlagen vor, koaktive Learning als ein Modell der Wechselwirkung des Dialogs zwischen einem Lernsystem und menschlichen Benutzer, wobei sowohl die beide das gemeinsame Ziel haben die Ergebnisse der maximalen Nutzen für den Benutzer zu liefern.

Coactive Preference Perceptron

- Model
 - Linear model of user utility: $U(y|x) = w^T \phi(x,y)$
- Algorithm
 - FOR $t = 1$ TO T DO
 - Observe x_t
 - Present $y_t = \operatorname{argmax}_y \{ w_t^T \phi(x_t, y) \}$
 - Obtain feedback \tilde{y}_t from user
 - Update $w_{t+1} = w_t + \phi(x_t, \tilde{y}_t) - \phi(x_t, y_t)$
- This may look similar to a multi-class Perceptron, but
 - Feedback \tilde{y}_t is different (not get the correct class label)
 - Regret is different (misclassifications vs. utility difference)

$$R(A) = \frac{1}{T} \sum_{t=1}^T [U(y_t^* | x_t) - U(y_t | x_t)]$$

[Shivaswamy, Joachims, 2012]

Coactive Perceptron: Regret Bound

- Model
 - $U(y|x) = w^T \phi(x,y)$, where w is unknown
- Feedback: ξ -Approximately α -Informative
 - $E[U(x_t, \tilde{y}_t)] \geq U(x_t, y_t) + \alpha(U(x_t, y_t^*) - U(x_t, y_t)) - \xi_t$
- Theorem
 - user feedback
 - system prediction
 - gap to optimal
 - model error

For user feedback \tilde{y} that is α -informative in expectation, the expected average regret of the Preference Perceptron is bounded by

$$E \left[\frac{1}{T} \sum_{t=1}^T U(y_t^* | x_t) - U(y_t | x_t) \right] \leq \frac{1}{\alpha T} \sum_{t=1}^T \xi_t + \frac{2R \|w\|}{\alpha \sqrt{T}}$$

[Shivaswamy, Joachims, 2012]

Preference Perceptron: Experiment

Experiment:

- Automatically optimize Arxiv.org Fulltext Search

Model

- Utility of ranking y for query x : $U_t(y|x) = \sum_i \gamma_i w_i^T \phi(x, y^{(i)})$ [~ 1000 features]
 \rightarrow Computing argmax ranking: sort by $w_i^T \phi(x, y^{(i)})$

Feedback

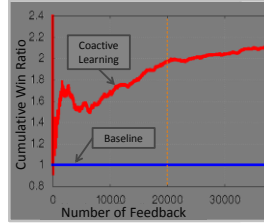
- Construct \tilde{y}_t from y_t by moving clicked links one position higher.
- Perturbation [Raman et al., 2013]

Baseline

- Handtuned w_{base} for $U_{\text{base}}(y|x)$

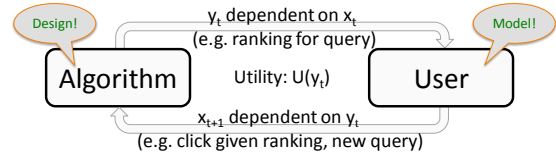
Evaluation

- Interleaving of ranking from $U_t(y|x)$ and $U_{\text{base}}(y|x)$



[Raman et al., 2013]

Interactive Learning System

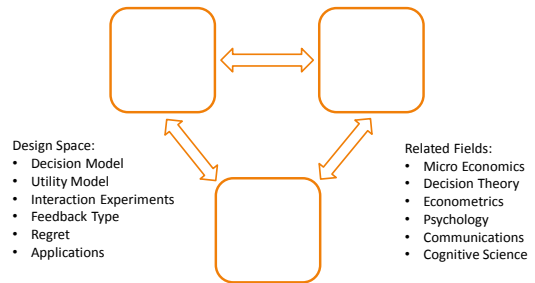


- Information Elicitation Interventions
- Decisions \rightarrow Feedback \rightarrow Learning Algorithm
 - Dueling Bandits
 - Model: Pairwise comparison test $P(y_i > y_j | U(y_i) > U(y_j))$
 - Algorithm: Interleaved Filter 2, $O(|Y| \log(T))$ regret
 - Coactive Learning
 - Model: for given y , user provides \tilde{y} with $U(\tilde{y}|x) > U(y|x)$
 - Algorithm: Preference Perceptron, $O(\|w\| T^{0.5})$ regret

Running Interactive Learning Experiments

- ~~Build your own system and provide service~~
 \rightarrow a lot of work
 \rightarrow too little data
- ~~Convince others to run your experiments on commercial system~~
 \rightarrow good luck with that
- Use large-scale historical log data from commercial system

Learning from Human Decisions



Contact: tj@cs.cornell.edu
 Software + Papers: www.joachims.org